Branching Distributional Equations and their Applications

Mariana Olvera-Cravioto

UNC Chapel Hill

molvera@unc.edu

August 22nd, 2018

Google's PageRank

- PageRank computes the rank of a webpage as:

$$r_i = (1-c)q_i + c\sum_{j\to i}\frac{r_j}{D_j^-},$$

where, $\{1, 2, \ldots, n\}$ are the pages under consideration, the sum is taken over all pages pointing to i, D_j^- is the number of outbound links of page j, $\mathbf{q} = (q_1, \ldots, q_n)$ is a personalization vector, and c is a damping factor, usually $c = 0.85$.

- Multiply both sides by n to obtain a "scale free" rank.
- In matrix notation,

$$\mathbf{R} = (1-c)\mathbf{Q} + \mathbf{RM}, \qquad \mathbf{M} = \text{matrix of weights}$$
$$\mathbf{Q} = \text{personalization vector.}$$

The matrix of weights

- Matrix $\mathbf{M}$ has elements:

$$M_{ij} = \frac{c}{D_i^-} A_{ij} 1(D_i^- \geq 1) = \frac{c}{D_i^- \vee 1} A_{ij},$$

  where

$$A_{ij} = \#\text{ arcs from } i \text{ to } j$$

- Matrix $\mathbf{A} = (A_{ij})$ is the **adjacency** matrix of the graph.
- **Note:**

$$\sum_{i=1}^{n} A_{ij} = D_j^+ \qquad \text{and} \qquad \sum_{j=1}^{n} A_{ij} = D_i^-$$

- Let $\mathbf{\Lambda} = \text{diag}(D_1^- \vee 1, \ldots D_n^- \vee 1)$, then

$$\mathbf{M} = c\mathbf{\Lambda}^{-1}\mathbf{A} \qquad \text{and} \qquad \|\mathbf{M}\|_\infty = \max_{1 \leq i \leq n} \sum_{j=1}^{n} M_{ij} = c < 1$$

Matrix iterations

- For any directed graph:

$$\mathbf{R} = (1-c)\mathbf{Q} + \mathbf{R}\mathbf{M} \quad \text{or equivalently,} \quad \mathbf{R} = (\mathbf{I} - \mathbf{M})^{-1}(1-c)\mathbf{Q}.$$

- Since $\mathbf{M}^k \to 0$ as $k \to \infty$, $\mathbf{R}$ has the representation

$$\mathbf{R} = (1-c)\mathbf{Q}\sum_{i=0}^{\infty}\mathbf{M}^i.$$

- Hence, we can approximate $\mathbf{R}$ with finitely many matrix iterations

$$\mathbf{R}^{(n,k)} := (1-c)\mathbf{Q}\sum_{i=0}^{k}\mathbf{M}^i.$$

Matrix iterations

- For any directed graph:

$$\mathbf{R} = (1-c)\mathbf{Q} + \mathbf{R}\mathbf{M} \quad \text{or equivalently,} \quad \mathbf{R} = (\mathbf{I} - \mathbf{M})^{-1}(1-c)\mathbf{Q}.$$

- Since $\mathbf{M}^k \to 0$ as $k \to \infty$, $\mathbf{R}$ has the representation

$$\mathbf{R} = (1-c)\mathbf{Q}\sum_{i=0}^{\infty}\mathbf{M}^i.$$

- Hence, we can approximate $\mathbf{R}$ with finitely many matrix iterations

$$\mathbf{R}^{(n,k)} := (1-c)\mathbf{Q}\sum_{i=0}^{k}\mathbf{M}^i.$$

- **Remark:** $\mathbf{R}^{(k)}$ contains only the "local" behavior of the graph.

Ignoring the tail of the series

- We want to bound $\|\mathbf{R}^{(n,\infty)} - \mathbf{R}^{(n,k)}\|_1$, where $\mathbf{R}^{(n,\infty)} := \mathbf{R}$.

Ignoring the tail of the series

- We want to bound $\|\mathbf{R}^{(n,\infty)} - \mathbf{R}^{(n,k)}\|_1$, where $\mathbf{R}^{(n,\infty)} := \mathbf{R}$.
- Note that:

$$\|\mathbf{R}^{(n,\infty)} - \mathbf{R}^{(n,k)}\|_1 = \left\|(1-c)\mathbf{Q}\sum_{r=k+1}^{\infty}\mathbf{M}^r\right\|_1 \leq (1-c)\sum_{r=k+1}^{\infty}\|\mathbf{Q}\mathbf{M}^r\|_1$$

$$\leq (1-c)\sum_{r=k+1}^{\infty}\|\mathbf{Q}\|_1\,\|\mathbf{M}^r\|_\infty$$

$$\leq (1-c)\sum_{r=k+1}^{\infty}\|\mathbf{Q}\|_1\,\|\mathbf{M}\|_\infty^r$$

$$= (1-c)\,\|\mathbf{Q}\|_1\sum_{r=k+1}^{\infty}c^r = c^{k+1}\,\|\mathbf{Q}\|_1$$

Exploring a randomly chosen vertex

- Consider either a DCM or an IRD on n vertices.
- Let $\mathbb{P}_n(\cdot)$ denote the conditional probability given the bi-degree sequence in the DCM or the weight sequence in the IRD.
- Let ξ be uniformly distributed in $\{1, 2, \ldots, n\}$.
- We want to approximate $R_\xi^{(n,\infty)}$ with $R_\xi^{(n,k)}$.

Exploring a randomly chosen vertex

- Consider either a DCM or an IRD on n vertices.
- Let $\mathbb{P}_n(\cdot)$ denote the conditional probability given the bi-degree sequence in the DCM or the weight sequence in the IRD.
- Let ξ be uniformly distributed in $\{1, 2, \ldots, n\}$.
- We want to approximate $R_\xi^{(n,\infty)}$ with $R_\xi^{(n,k)}$.
- Note that:

$$\begin{aligned}
\mathbb{P}_n\left(\left|R_\xi^{(n,k)} - R_\xi^{(n,\infty)}\right| > \epsilon\right) &\leq \frac{1}{\epsilon}\mathbb{E}_n\left[\left|R_\xi^{(n,k)} - R_\xi^{(n,\infty)}\right|\right] \\
&= \frac{1}{\epsilon n}\sum_{i=1}^{n}\mathbb{E}_n\left[\left|R_i^{(n,k)} - R_i^{(n,\infty)}\right|\right] \\
&= \frac{1}{\epsilon n}\mathbb{E}_n\left[\left\|\mathbf{R}^{(n,\infty)} - \mathbf{R}^{(n,k)}\right\|_1\right] \\
&\leq \frac{c^{k+1}}{\epsilon n}\left\|\mathbf{Q}\right\|_1 .
\end{aligned}$$

Approximation with the local neighborhood

- The usual assumptions imply that $n^{-1}\|\mathbf{Q}\|_1 = O(1)$ as $n \to \infty$.
- It follows that

$$\mathbb{P}_n\left(\left|R_\xi^{(n,k)} - R_\xi^{(n,\infty)}\right| > \epsilon\right) = O(\epsilon^{-1}c^{k+1}).$$

- We usually choose $k = \delta \log n$ and $\epsilon = O(n^{-\eta})$ for carefully chosen $\delta, \eta > 0$.
- **Next step:** Approximate $R_\xi^{(n,k)}$ by the PageRank computed on a marked branching process up to its kth generation.

Coupling with a suitable branching process

- Choose randomly one vertex in the graph and perform a breadth-first exploration of its in-component.
- **DCM:**
  - Couple with a marked Galton-Watson process.
  - The marks are $\{(D_i^-, Q_i) : 1 \leq i \leq n\}$.
  - Root has distribution

$$f_n^*(i,j,t) = \frac{1}{n}\sum_{k=1}^{n} 1(D_k^+ = i, D_k^- = j, Q_k = t).$$

  - All other nodes have distribution

$$f_n(i,j,t) = \sum_{k=1}^{n} 1(D_k^+ = i, D_k^- = j, Q_k = t)\frac{D_k^-}{L_n},$$

where $L_n = \sum_{i=1}^{n} D_k^-$.

Coupling with a suitable branching process... cont.

- **IRD:**
  - Couple with a marked multi-type BP with types $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ sampled from μ.
  - Types are of the form $\mathbf{x} = (x^+, x^-, q)$, marks are of the form D^-.
  - Root has type distributed according to:

$$f_n^*(\mathbf{x}) = \frac{1}{n} \sum_{k=1}^{n} 1(\mathbf{x}_i = \mathbf{x})$$

  - The number of offspring of type $\mathbf{x}_j$ that an individual of type $\mathbf{x}_i$ has is Poisson with mean:

$$\frac{\kappa(\mathbf{x}_j, \mathbf{x}_i)}{\theta n},$$

    where $\theta = \lim_{n\to\infty} n^{-1} \sum_{k=1}^{n} (X_i^+ + X_i^-)$.
  - Observation, for rank-1 kernels, i.e., $\kappa(\mathbf{x}, \mathbf{y}) = \kappa_-(\mathbf{x})\kappa_+(\mathbf{y})$ we have that the type of an offspring is independent of the type of its parent (Exercise).
  - For rank-1 kernels, the multi-type BP can be seen as a Galton-Watson process.

Coupling theorem

- For each model we construct couplings between a breadth-first exploration of the in-component of a randomly chosen vertex and either:
  - a marked Galton-Watson process for the DCM, or
  - a marked multi-type BP for the IRD.
- We say that the coupling **holds** up to a depth k if the two processes are "identical" up to trees of depth k.
- **Theorem:** There exists $\omega > 0$ such that the coupling holds with high probability up to a depth $k_n = \omega \log n$.
- Let $\hat{R}_{\emptyset}^{(n,k)}$ denote the PageRank of the root node of the coupled tree computed up to generation k.
- Note that when the coupling holds up to k_n we have:

$$R_{\xi}^{(n,k_n)} = \hat{R}_{\emptyset}^{(n,k_n)}$$

Convergence to a limiting tree

- The coupling trees in both cases have distributions that still depend on either the bi-degree or type sequence of a graph on n vertices.
- The next step is to show that

$$\hat{R}_{\emptyset}^{(n,k_n)} \Rightarrow \mathcal{R}^*$$

as $n \to \infty$, where $\mathcal{R}$ corresponds to the PageRank of the root of a limiting marked BP.

- In particular,

$$\mathcal{R}^* = \sum_{\mathbf{i} \in \mathcal{T}} \Pi_{\mathbf{i}} \mathcal{Q}_{\mathbf{i}},$$

where the $\Pi_{\mathbf{i}}$ are constructed on a weighted branching process whose root has a different distribution than all other nodes.

The main theorems

- Consider either the DCM or the IRD.
- Let $R_\xi^{(n,\infty)}$ denote the PageRank of a randomly chosen vertex.
- **Theorem:** Under some technical conditions on the bi-degree or type sequences, we have that

$$R_\xi^{(n,\infty)} \Rightarrow \mathcal{R}^*, \qquad n \to \infty,$$

where

$$\mathcal{R}^* = (1-c)\mathcal{Q}_0 + \sum_{j=1}^{\mathcal{N}_0} X_j,$$

with the $\{X_j\}$ are i.i.d. copies of the attracting endogenous solution to the linear distributional equation:

$$X \stackrel{\mathcal{D}}{=} (1-c)c\frac{\mathcal{Q}}{\mathcal{D}} + \sum_{j=1}^{\mathcal{N}} \frac{c}{\mathcal{D}} X_j,$$

and $(\mathcal{N}, \mathcal{D})$ the limiting size-biased in-degree and out-degree of the graph.